

# DNA-binding sequence of the human prostate-specific homeodomain protein NKX3.1

David J. Steadman, Domenica Giuffrida and Edward P. Gelmann\*

Department of Oncology, Lombardi Cancer Center, Georgetown University School of Medicine,  
3800 Reservoir Road NW, Washington, DC 20007-2197, USA

Received February 2, 2000; Revised and Accepted April 21, 2000

## ABSTRACT

**NKX3.1 is a member of the NK class of homeodomain proteins and is most closely related to *Drosophila* NK-3. NKX3.1 has predominantly prostate-specific expression in the adult human. Previous studies suggested that NKX3.1 exerts a growth-suppressive effect on prostatic epithelial cells and controls differentiated glandular functions. Using a binding site selection assay with recombinant NKX3.1 protein we identified a TAAGTA consensus binding sequence that has not been reported for any other NK class homeoprotein. By electromobility shift assay we demonstrated that NKX3.1 preferentially binds the TAAGTA sequence rather than the binding site for Nkx2.1 (CAAGTG) or Msx1 (TAATTG). Using mutated binding sites in competitive gel shift assays, we analyzed the nucleotides in the TAAGTA consensus sequence that are important for NKX3.1 binding. The consensus binding site of a naturally occurring polymorphic NKX3.1 protein with arginine replaced by cysteine at position 52 was identical to the wild-type binding sequence. The binding affinities of wild-type and polymorphic NKX3.1 for the TAAGTA consensus site were very similar, with values of 20 and 22 nM, respectively. Wild-type and polymorphic NKX3.1 specifically repressed transcription of luciferase from a reporter vector with three copies of the NKX3.1-binding site upstream from a thymidine kinase promoter. The data show that among NK family proteins NKX3.1 binds a novel DNA sequence and can behave as an *in vitro* transcriptional repressor.**

## INTRODUCTION

Homeobox genes encode proteins that contain highly conserved DNA-binding regions called homeodomains. Homeodomain proteins, or homeoproteins, bind to specific DNA sequences and function as transcription factors that regulate eukaryotic development with spatial and temporal specificity (1). NKX3.1 is a novel homeoprotein that belongs to the *Drosophila* NK class of homeodomain proteins. NKX3.1 is most closely related to *Drosophila* NK-3 by virtue

of 78% sequence homology to the NK-3 homeodomain region (2–4). Murine Nkx3.1 has 100% sequence homology to the human NKX3.1 homeodomain region and shares 90% sequence homology overall (3–6).

Expression of murine *Nkx3.1* in the developing embryo and fetus has been detected in a variety of tissue types, including mesoderm, vascular smooth muscle, epithelium and regions of the central nervous system (7–9). However, in the adult mouse *Nkx3.1* expression is androgen regulated (5) and restricted primarily to the prostate and the bulbourethral gland (6,9). *NKX3.1* expression in the adult human is also androgen regulated and localized predominantly in the prostate, with low levels also detected in the testis (3,4).

*NKX3.1* has been mapped to human chromosome 8p21, a locus that is frequently deleted in prostate cancer (10–13). Therefore, *NKX3.1* is a candidate tumor suppressor gene. However, no cancer-specific mutations of the *NKX3.1* coding region were found in human prostate cancer samples (10). Nevertheless, the potential for NKX3.1 to exert a differentiating and growth suppressing effect on prostatic epithelium was confirmed by targeted gene disruption of *Nkx3.1* in mice (9). Deletion of either one or both copies of *Nkx3.1* resulted in prostatic epithelial hyperplasia and dysplasia that increased in severity with age. Homozygous deletion of *Nkx3.1* caused defective prostate ductal morphogenesis and decreased seminal protein production. Loss of NKX3.1 protein expression may be important in prostate cancer pathogenesis. We have recently shown that ~40% of human prostate cancer samples had diminished expression of NKX3.1 compared to adjacent normal epithelium (Bowen *et al.*, submitted for publication).

Genetic analysis indicated that ~10–15% of human DNA samples contain a polymorphic *NKX3.1* gene characterized by the presence of a C→T polymorphism at nucleotide 154 (10; our unpublished data). The polymorphism resulted in the substitution of arginine by cysteine at codon 52, located N-terminal to the homeodomain. To date, a phenotype has not been identified for the polymorphic protein.

NK-2 class homeoproteins preferentially bind DNA sequences with a CAAG core sequence (14,15), while other NK class homeoproteins preferentially bind TAAT core sequences (16,17). The optimal DNA binding site for *Drosophila* NK-3 or its mammalian homologs has yet to be defined. Using a binding site selection assay, we identified the *in vitro* DNA binding site of NKX3.1. Electromobility gel shift assays were used to analyze binding affinity and nucleotides in the consensus site that are important for NKX3.1 binding. By

\*To whom correspondence should be addressed. Tel: +1 202 687 2207; Fax: +1 202 784 1229; Email: gelmanne@gunet.georgetown.edu

inserting three copies of the NKX3.1-binding site into a luciferase reporter vector, we investigated the transcriptional function of NKX3.1 using reporter gene analysis. Finally, the DNA binding and transcriptional activities of wild-type and polymorphic (R52C) NKX3.1 were compared to determine if differences existed between the proteins.

## MATERIALS AND METHODS

### Expression and reporter plasmid construction

Full-length NKX3.1 cDNA was synthesized by PCR amplification of full-length wild-type human NKX3.1 cDNA obtained from a normal prostate cDNA library (3). Amplification was carried out for 30 cycles (94°C for 30 s, 62°C for 30 s and 72°C for 60 s) using a Perkin Elmer temperature cyclor (Perkin Elmer, Norwalk, CT). The primers (Gibco BRL, Rockville, MD) used in the amplification incorporated *EcoRI* and *XhoI* restriction sites for directional cloning of the amplified DNA into expression vectors. The forward PCR primer has the sequence 5'-CGG-GATCCGAATTCATGCTCAGGGTTCCGGAGCCGC-3' and the reverse primer has the sequence 5'-GGGCCTCGAGT-CTAGAGTTACCCAAAAGCTGGGCTCCA-3'. The amplified DNA was cloned into a pCRII-Topo vector (Invitrogen, Carlsbad, CA), following the protocol provided by the manufacturer. The DNA was digested with *EcoRI* and *XhoI* restriction enzymes and the fragment representing NKX3.1 cDNA was excised from a 1% agarose gel. The DNA was purified using a QiaexII DNA isolation kit (Qiagen, Valencia, CA) and the resulting fragment was directionally cloned into the *EcoRI* and *XhoI* restriction sites of plasmid pcDNA3 (Invitrogen) for mammalian protein expression. NKX3.1 cDNA was also cloned into the *EcoRI* and *Sall* sites of the pMAL-C2G vector (New England Biolabs, Beverly, MA) for bacterial protein expression. Wild-type NKX3.1 cDNA was altered using a Stratagene Quikchange Site-Directed Mutagenesis kit following the manufacturer's protocol (Stratagene, La Jolla, CA). R52C NKX3.1 was generated by mutating the nucleotide sequence of codon 52 of wild-type NKX3.1 cDNA from CGC to TGC, thereby changing the corresponding amino acid from arginine to cysteine. The mutant NKX3.1 cDNA was sequenced entirely to ensure that no additional mutations were generated.

Luciferase reporter vectors were constructed by ligating an insert with partial *XhoI* ends containing three copies of TAAGTA or CACGTG into the *XhoI* site of pT109, which contains a herpes simplex virus thymidine kinase (TK) promoter upstream of a firefly luciferase gene. The sequence of one strand of the TAAGTA insert (binding sites underlined) containing a partial *XhoI* site was 5'-TCGATATTAAGTAT-AGGATTAAGTATAGGGATTAAGTAT-3' and the sequence of one strand of the CACGTG insert was 5'-TCG-ATATCACGTGTAGGATCACGTGTAGGGATCACGTGT-3'. The orientation of the TAAGTA insert was confirmed by sequence analysis.

### Bacterial protein expression and purification of recombinant NKX3.1

Plasmids containing wild-type or R52C NKX3.1 cDNA inserted into pMAL-C2G were used to transform competent *Escherichia coli* strain BL21 using standard techniques.

Transformants were used to inoculate 10 ml of Luria Bertani (LB) medium and the cultures were incubated with shaking at 220 r.p.m. overnight at 37°C. This culture was used to inoculate 1 l of LB medium containing 0.2% glucose, followed by incubation at 37°C with shaking at 220 r.p.m. until the optical density at 600 nm reached 0.6. Isopropyl- $\beta$ -D-thiogalactopyranoside was added to a final concentration of 0.3 mM and the culture was incubated for an additional 60 min. Cells were harvested by centrifugation and frozen at -80°C. Approximately 1 g of frozen cells were thawed on ice and resuspended in 10 ml of buffer A [20 mM Tris-HCl, 200 mM NaCl, 1 mM EDTA and 1 mM dithiothreitol (DTT)]. Suspended cells were lysed by sonication and a cell-free extract was prepared by centrifugation. The cell-free extract was loaded onto an amylose column (New England Biolabs) and washed with buffer A. NKX3.1 fusion proteins were eluted from the column with buffer A containing 10 mM maltose. Protein purity was demonstrated by 12% SDS-PAGE and purified protein concentration was determined by the Bradford assay (18).

### Selection and amplification binding assay

A selection and amplification binding (SAAB) assay was performed essentially as described for murine Nkx2.5 (15), with modifications. Briefly, 0.5 pmol of a 15 bp random sequence flanked by 20 bp regions of non-random sequence was radiolabeled with 10  $\mu$ Ci of [ $\gamma$ -<sup>32</sup>P]ATP using T4 polynucleotide kinase. The radiolabeled probe was incubated with 25 pmol of purified wild-type or R52C NKX3.1 fusion protein in binding buffer [10 mM Tris-HCl, pH 7.5, 50 mM NaCl, 7.5 mM MgCl<sub>2</sub>, 1 mM EDTA, 5% glycerol, 5% sucrose, 0.1% Nonidet P-40, 0.1  $\mu$ g BSA, 1  $\mu$ g poly(dI-dC) and 5 mM DTT]. Unbound probe was separated from protein-bound DNA by native 8% PAGE in 0.5× TBE buffer. Protein-bound DNA was detected by autoradiography and the bands representing protein-DNA complexes were excised from the gel. The DNA was eluted from the gel slice overnight at 37°C in elution buffer (0.5 M ammonium acetate, 1 mM EDTA and 0.1% SDS). Eluted DNA was PCR amplified using primers complementary to the 20 bp non-random flanking sequences and purified by native 12% PAGE in 1× TBE buffer. The purified DNA was radiolabeled and used as a probe for another round of SAAB for a total of five rounds. The amount of protein was reduced to 5 pmol for the fourth and fifth rounds of SAAB. Following the final PCR amplification and gel purification, the 55 bp fragment was cloned into pCRII-TOPO using a TOPO-TA cloning kit (Stratagene). The inserts were sequenced on an ABI 377 Nucleotide Sequencer (Perkin Elmer).

### Electromobility gel shift assay

Double-stranded DNA used as probe or competitor was prepared by annealing equimolar amounts of complementary oligonucleotides with flanking partial *BamHI* ends and filled-in with Klenow polymerase. The double-stranded DNA used as probe was radiolabeled with [ $\gamma$ -<sup>32</sup>P]ATP using T4 polynucleotide kinase. The labeled probe (1 pmol) was added to wild-type or R52C NKX3.1 (125 ng) in binding buffer to a final volume of 20  $\mu$ l. Competition binding experiments were performed essentially as described by Amendt *et al.* (19) by including unlabeled competitor DNA concentrations of 50, 250 and 500 nM. Competitor DNAs were preincubated with protein at room temperature for 30 min prior to addition of radiolabeled

probe. The nucleotide sequence of one competitor strand is listed, where used, in the appropriate figure. For dissociation constant ( $K_d$ ) measurements,  $1 \times 10^{-10}$  M radiolabeled probe was incubated with protein concentrations ranging from 0.5 to  $200 \times 10^{-9}$  M for 1 h prior to electrophoresis. Protein–DNA complexes were separated from unbound probe by native 8% PAGE at 175 V for 1 h at ambient temperature in  $0.5 \times$  TBE buffer following pre-electrophoresis of the gel at 200 V for 60 min. Following electrophoresis, a Molecular Dynamics PhosphorImager Screen (Molecular Dynamics, Sunnyvale, CA) was exposed to the dried gel. A Molecular Dynamics PhosphorImager and ImageQuaNT software (Molecular Dynamics) were used in the quantitative analyses. Binding competition was determined by calculating the ratio of bound to free probe normalized to the absence of competitor DNA. Dissociation constants were calculated from the binding data by plotting bound protein versus free protein and using the expression  $K_d = [D][P]/[DP]$  in a non-linear least squares analysis of the data. Free DNA is represented by [D], [P] is the free protein concentration and [DP] is the concentration of DNA–protein complex. Competition binding and dissociation constant measurements were performed at least three times and representative gel shift assays from the competition binding experiments are shown.

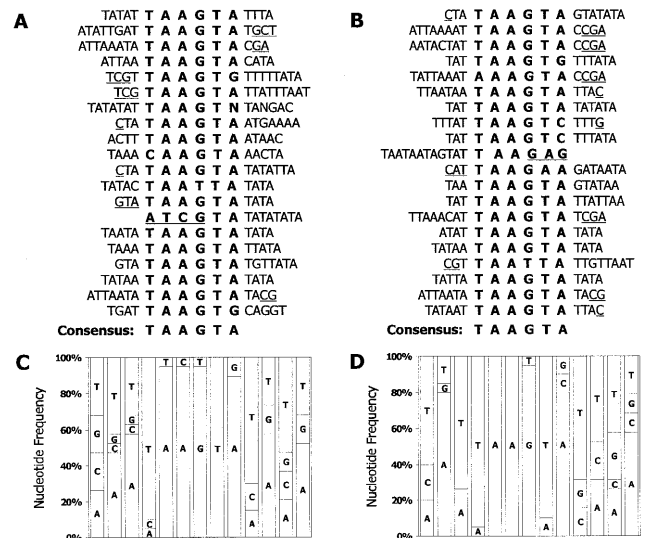
### Cell culture and reporter gene assays

The prostate carcinoma cell line TSU-Prl (20) was cultured in Improved Minimal Essential Medium (Gibco BRL) supplemented with 5% fetal calf serum. Cells were transiently transfected using  $6 \times 10^4$  cells/well in a 24-well plate with 250 ng reporter plasmid, 250 ng expression plasmid and 1 ng CMV-*Renilla* plasmid in the presence of 2  $\mu$ g LipofectAMINE 2000 (Gibco BRL) following the manufacturer's protocol. Cells were lysed 48 h after transfection and the lysate was assayed for firefly and *Renilla* luciferase activities using Dual Luciferase Reporter Assay reagents (Promega, Madison, WI). The data represent the average of three separate experiments normalized by *Renilla* luciferase activity.

## RESULTS

### Determination of the NKX3.1 consensus binding sequence by SAAB assay

The SAAB assay originally described by Blackwell and Weintraub (21) with modifications by Chen and Schwartz (15) was used to determine a consensus DNA binding sequence for bacterially expressed wild-type and R52C NKX3.1 fusion proteins. Purified proteins were used to select preferred binding sequences from a random pool of double-stranded DNA. Five cycles of binding selection were used to isolate optimal binding sequences and the enriched double-stranded DNA was cloned and sequenced. Alignment of the 20 selected sequences each for wild-type (Fig. 1A) and R52C NKX3.1 (Fig. 1B) revealed an identical consensus sequence of 6 nt (TAAGTA) with an individual nucleotide frequency of occurrence ranging from 78 to 100% (Fig. 1C and D). The NKX3.1 consensus sequence is similar, but not identical, to the binding sequences of other NK-2 class homeoproteins such as the CAAGTG site for Nkx2.1 (14) and TNAAGTG for Nkx2.5 (15). The consensus sequence for NKX3.1 is also similar to the TAAGTG consensus binding



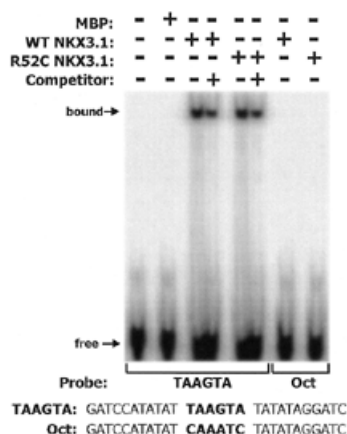
**Figure 1.** Determination of the NKX3.1 consensus DNA binding sequence. Sequences selected by (A) recombinant wild-type or (B) R52C NKX3.1 following five rounds of SAAB are listed. Underlined nucleotides represent non-random sequence. Nucleotides shown in bold represent those used in determination of the consensus binding sequence. The overall consensus binding sequence for each protein is listed below the aligned sequences. The percentage frequencies of each nucleotide for sequences selected by (C) wild-type NKX3.1 and (D) R52C NKX3.1 are shown.

sites of human Hox11 (22) and salmon Isl-2 (23), members of the Hox and LIM families of homeoproteins, respectively. The selected sequences also contained the TAAT motif recognized by a majority of homeodomain proteins with a characterized DNA-binding site (1). The TAAT motif was present at least once in 88% of the combined wild-type and R52C selected sequences. In similar binding assays Nkx2.5 and Isl-2 also selected sequences containing TAAT, in addition to their consensus binding sites (15,23).

Specific binding of wild-type and R52C NKX3.1 to the TAAGTA site was demonstrated by gel mobility shift (Fig. 2). Maltose-binding protein lacking the NKX3.1 moiety did not bind the TAAGTA sequence, suggesting that protein–DNA interactions were specifically due to the NKX3.1 segment of the fusion protein. Binding specificity was confirmed by competition with a 5-fold molar excess of unlabeled NKX3.1 consensus sequence, which competed for the TAAGTA probe. An Oct POU-1 homeodomain binding site was included as a control for non-specific DNA binding. The Oct site contains a CTAAAC binding site rather than the preferentially bound TAAGTA sequence. Neither protein bound the Oct probe.

### Comparison of wild-type NKX3.1 binding to the NKX3.1, Nkx2.1 and Msx1 binding sites

The results of the SAAB assay indicated that NKX3.1 preferentially bound a TAAGTA sequence. Using a competitive binding assay, we compared wild-type NKX3.1 binding to the TAAGTA site with binding to the Nkx2.1 and Msx1 DNA binding sites. The Nkx2.1 sequence (14) has a CAAG core

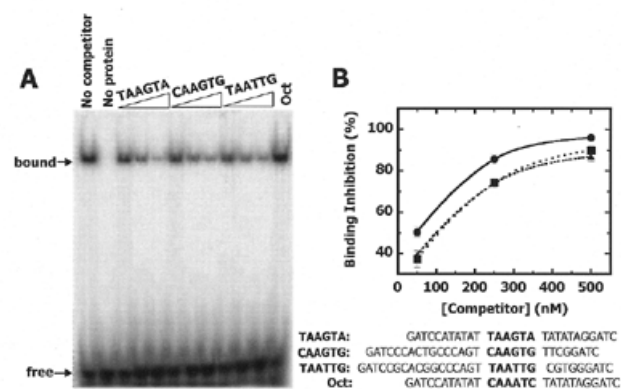


**Figure 2.** NKX3.1 specifically binds to the TAAGTA consensus binding site. Binding of maltose-binding protein, wild-type NKX3.1 and R52C NKX3.1 to a radiolabeled probe containing either the NKX3.1 consensus sequence (TAAGTA) or an Oct-1 binding site (CTAAAC) was analyzed by gel mobility shift. Free probe and protein-bound complex are indicated with arrows. Where indicated, a 5-fold molar excess of unlabeled TAAGTA sequence was included as competitor. Nucleotide sequences of the probes and competitor are listed at the bottom of the figure.

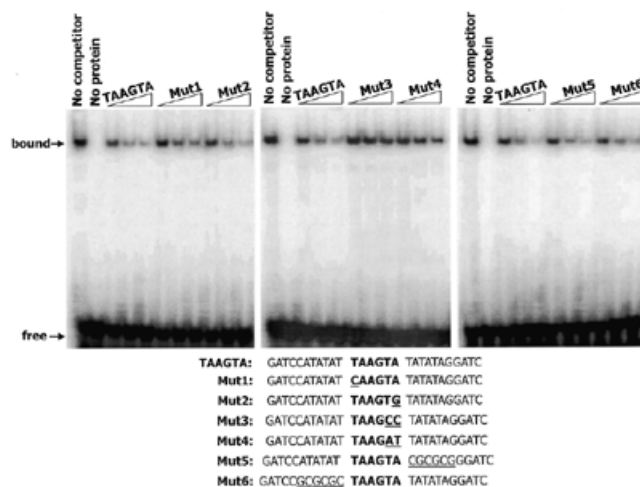
binding site, while the Msx1-binding site (24) contains the TAAT core sequence recognized by most homeoproteins. The nucleotide sequences of the Nkx2.1 (CAAGTG) and the Msx1 (TAATTG) binding sites are similar to the NKX3.1 consensus sequence, but differ at nucleotides proposed to be important for optimal binding (1,25). The results, shown in Figure 3A, suggest that wild-type NKX3.1 will bind to the Nkx2.1 and Msx1 consensus sites. These results are in agreement with a previous report that showed that a murine Nkx3.1 homeodomain polypeptide bound both CAAGTG and TAATTG sequences (5). However, the present data indicate that the NKX3.1 TAAGTA consensus site was reproducibly a slightly stronger competitor for NKX3.1 binding than either the CAAGTG or TAATTG binding sites (Fig. 3B). Using increasing amounts of protein with a constant probe concentration, Chen *et al.* also observed only minor differences in the binding affinity of Nkx2.5 when the CAAGTG NK-2 consensus sequence was substituted with either TAATTA or TAAGTG (15). As a control for non-specific competitor binding, an Oct-1 CTAAAC binding site, which was not bound by NKX3.1 (Fig. 1), was included as a competitor at 500 nM, the highest concentration used in the binding assay. There was no apparent decrease in the band representing protein-bound DNA in the presence of CTAAAC.

#### Effect of mutations in the TAAGTA consensus site on NKX3.1 binding

NKX3.1 belongs to the NK class of homeoproteins, but preferentially bound the novel TAAGTA site rather than the Nkx2.1 CAAGTG site. Competitive gel shift experiments were used to isolate the nucleotide(s) important for preferential binding of NKX3.1 to the TAAGTA site. Sequences were generated to replace either T with C at position 1 (TAAGTA→CAAGTA)

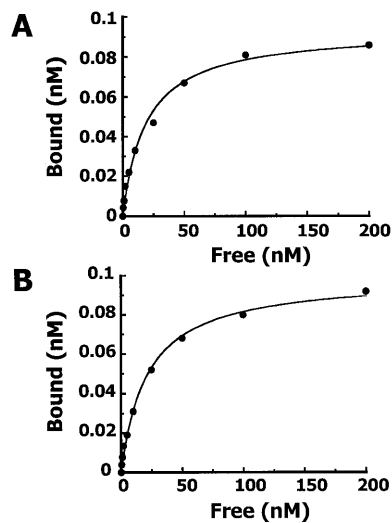


**Figure 3.** Wild-type NKX3.1 preferentially binds the NKX3.1 consensus sequence rather than Nkx2.1 and Msx1 binding sites. (A) Binding affinity of NKX3.1 to TAAGTA was compared with NKX3.1 binding to the Nkx2.1 and Msx1 binding sites by a competitive gel shift assay. Wild-type NKX3.1 was incubated with a probe containing the NKX3.1 consensus sequence in the absence or presence of unlabeled competitor DNA. Competitors were included at concentrations of 50, 250 and 500 nM. Free probe and protein-bound complex are indicated with arrows. (B) The data were quantitated and normalized to wild-type NKX3.1 binding to probe with no competitor. Nucleotide sequences of the competitors are listed at the bottom of the graph. closed circles, TAAGTA; closed squares, CAAGTG; closed triangles, TAATTG.



**Figure 4.** Effect of mutations in the TAAGTA consensus site on NKX3.1 binding. The effects of mutations in the NKX3.1 consensus site were analyzed by competitive gel shift assay. Wild-type NKX3.1 was incubated with a probe containing the NKX3.1 consensus sequence in the absence or presence of unlabeled competitor DNA. Competitors were included at concentrations of 50, 250 and 500 nM. Free probe and protein-bound complex are indicated with arrows. Nucleotide sequences of the competitors are listed at the bottom of the figure.

or A with G at position 6 (TAAGTA→TAAGTG) of the NKX3.1 consensus binding site. The results of the competition assay (Fig. 4) indicate that replacing A at position 6 of the consensus site with G did not affect binding competition relative to the native TAAGTA site. However, replacing T with C at position 1 decreased binding competition, suggesting

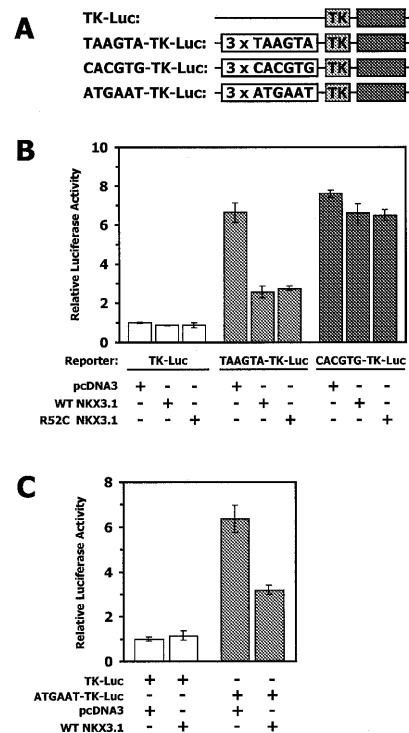


**Figure 5.** Wild-type and R52C NKX3.1 exhibit similar DNA binding affinities. The equilibrium dissociation constants for (A) wild-type and (B) R52C NKX3.1 binding to an NKX3.1 consensus site were determined by gel mobility shift using a constant amount of radiolabeled probe ( $1 \times 10^{-10}$  M) with various protein concentrations ( $0.5$ – $200 \times 10^{-9}$  M). A plot of the quantitated data is shown as bound protein–DNA complex concentration as a function of free protein concentration. The data were analyzed by non-linear least squares as described in Materials and Methods.

that T instead of C at the first position of the consensus sequence is preferred for optimal NKX3.1 binding. Mutating the TAAGTA site to either TAAGCC (Mut3) or TAAGAT (Mut4) significantly decreased competitor binding (Fig. 4), indicating that the 3'-dinucleotide of the NKX3.1 consensus site is important for protein binding. These results are in agreement with previous findings by Damante *et al.* (25,26) indicating that the 3'-dinucleotide of the Nkx2.1 site (CAAGTG) is necessary for optimal protein binding. Sequences selected by wild-type and R52C NKX3.1 in the SAAB assay (Fig. 1) exhibited a high frequency of A/T-rich nucleotides flanking the TAAGTA binding site. To investigate the importance of the A/T flanking sequence for wild-type NKX3.1 binding, a competitive gel shift assay was performed using NKX3.1 binding sites with A/T flanking nucleotides replaced with G/C either upstream (Mut5) or downstream (Mut6) of the TAAGTA sequence (Fig. 4). Both Mut5 and Mut6 had similar competitor efficacy as the native NKX3.1 binding sequence, indicating that the predominance of A/T nucleotides flanking the TAAGTA site did not significantly influence NKX3.1 DNA binding. It is possible that the high frequency of A/T base pairs flanking the NKX3.1 consensus site was an artifact of the binding assay. In fact, previous selection assays using NK class homeoproteins also yielded sequences with a relatively high A/T content outside their respective consensus binding sites (15,17).

#### Wild-type and R52C NKX3.1 exhibit similar binding affinities

To determine if the polymorphism at position 52 affected DNA binding affinity, the dissociation constants for wild-type and



**Figure 6.** NKX3.1 represses reporter gene transcription. (A) Diagram representing reporter plasmids used to analyze the effect of NKX3.1 on reporter gene transcription. (B) Reporter plasmids were used to transiently transfect TSU-Pr1 cells co-transfected with wild-type or R52C NKX3.1 expression plasmids or empty expression plasmid (pcDNA3) as described in Materials and Methods. CMV-*Renilla* plasmid was included as a control for transfection efficiency. (C) TSU-Pr1 cells were co-transfected with expression plasmid and empty reporter vector or reporter vector containing NKX3.1 binding sites in the antisense direction. CMV-*Renilla* plasmid was included as a control for transfection efficiency.

R52C NKX3.1 binding to the TAAGTA consensus site were measured. Increasing amounts of protein were incubated with a constant amount of radiolabeled probe. Following quantitative analysis, the data were plotted and analyzed by non-linear least squares for dissociation constant determination (Fig. 5). The dissociation constants for the proteins were nearly identical, exhibiting  $K_d$  values of 20 nM for wild-type and 22 nM for R52C NKX3.1. The results indicate that mutation at position 52 of NKX3.1, which lies N-terminally outside the homeo-domain region, does not affect DNA binding affinity.

#### NKX3.1 acts as a transcriptional repressor

Luciferase reporter vectors (Fig. 6A) were constructed with three tandem NKX3.1-binding sites in the sense (TAAGTA-TK-Luc) and antisense (ATGAAT-TK-Luc) directions upstream from a herpes simplex virus TK promoter. Co-transfection of TSU-Pr1 cells with TAAGTA-TK-Luc and an NKX3.1 expression vector resulted in decreased transcription compared with control cells co-transfected with TAAGTA-TK-Luc and empty expression vector (Fig. 6B), suggesting that in this context NKX3.1 behaved as a transcriptional

repressor. There was no apparent difference in the relative luciferase activities of wild-type and R52C NKX3.1 co-transfected with TAAGTA-TK-Luc, suggesting that the polymorphism did not affect NKX3.1 transcriptional repression *in vitro*. Previous binding studies have shown that a murine Nkx3.1 homeodomain will not bind to a CACGTG sequence (5). When the NKX3.1-binding sites were mutated to CACGTG, the relative luciferase activity using a NKX3.1 expression plasmid had little effect on transcription, indicating that transcriptional repression with reporter constructs containing three copies of TAAGTA was due to specific binding of NKX3.1. The level of transcriptional repression using ATGAAT-TK-Luc and wild-type NKX3.1 expression vector was similar to that observed using TAAGTA-TK-Luc and wild-type expression vector (Fig. 6C), indicating that the orientation of the TAAGTA sequence does not affect transcriptional repression due to NKX3.1 binding. Insertion of the TAAGTA and CACGTG inserts increased luciferase activity relative to the unmodified TK-Luc reporter plasmid. Based on these results, it is likely that these inserts contain binding sites for transcriptional activators endogenously expressed in TSU-Pr1 cells. For example, the TAAGTA insert contains multiple ATTA sequences (the complement of TAAT), which is bound by most homeoproteins (1). Also, the CACGTG sequence is a high affinity MYC-binding site (27).

## DISCUSSION

Using purified fusion proteins in a random DNA selection assay, a TAAGTA consensus binding sequence for wild-type and R52C NKX3.1 was identified. The TAAGTA consensus site has not been isolated for any other NK class homeoprotein. Previously reported DNA binding sites for NK class homeoproteins include CAAGTG for Nkx2.1, AAAGTG for *Caenorhabditis elegans* ceh-22 and NAAGTG for Nkx2.5 (14,15,28). Structural and mutational studies indicate that amino acid residues of helix III and the N-terminal arm of the homeodomain primarily determine DNA binding specificity (26). Amino acid residues 6–8 of the N-terminal arms of the NK-2 and Nkx2.1 homeodomains have been implicated as the primary determinants of the preferred nucleotide for binding at the 5'-end of the 5'-CAAG-3' core binding site (26,29). The homeodomains of NK-2 and Nkx2.1 have Val, Leu and Phe at positions 6–8, while NKX3.1 has Ala, Ala and Phe at these positions. The variation in side-chain volume of residues 6 and 7 may result in preferential binding by NKX3.1 of T instead of C at the first position of the TAAGTA binding site. When Leu at position 7 of the NK-2 homeodomain was mutated to Ala, the binding affinity for the CAAGTG site decreased by one order of magnitude (29), further suggesting that side-chain volume of residues 6 and 7 is an important determinant of homeodomain binding specificity. Attempts to alter binding preference from TAAGTA to CAAGTA by mutation of Ala to Leu at position 7 of the NKX3.1 homeodomain were unsuccessful (our unpublished data). Therefore, it is likely that a relatively small side-chain volume at position 6, in addition to position 7, of the NKX3.1 homeodomain is required for preferential binding of T instead of C in the TAAGTA sequence. NK class homeodomains characteristically have Tyr at position 54 (2), which confers preference for atypical DNA binding sites

relative to the canonical TAAT binding site recognized by most homeodomains (26). The tyrosine residue at position 54 of the homeodomain is reportedly involved in the recognition and preferential binding of guanosine at the 3'-end of the 5'-CAAG-3' core binding site for Nkx2.1 (26). NKX3.1 also has Tyr at position 54 of the homeodomain, therefore, it is not surprising that wild-type NKX3.1 preferentially binds a TAAG rather than a TAAT core sequence (Fig. 3). Optimal DNA binding of Nkx2.1 requires a TG dinucleotide following the CAAG core binding site (25). The consensus DNA binding sequence for NKX3.1 exhibited a TA dinucleotide following the TAAG core. Structural studies of Nkx2.1 have revealed that Gln at position 50 of the homeodomain region contacts the TG dinucleotide flanking the 3'-end of the 5'-CAAG-3' core binding sequence. This residue is conserved among NK-2 and NK-3 homeodomains (2). Therefore, based on sequence homology, it is predicted that NKX3.1 would also preferentially bind a TG dinucleotide instead of TA following the TAAG core binding site. Although our binding site selection data suggest that NKX3.1 prefers a TA dinucleotide following the TAAG core (Fig. 4), competitive binding experiments suggested that replacing the A with G in the dinucleotide (TAAGTA→TAAGTG) had no observable effect on wild-type NKX3.1 binding (Fig. 4), providing evidence that the nucleotides may be interchangeable. In fact, a previous report describing the DNA sequences within the thyroglobulin promoter important for Nkx2.1 transactivation identified a CAAGTA binding site as well as CAAGTG (30).

Based on the results of competition gel shift assays, NKX3.1 displays a high degree of binding promiscuity (Fig. 3). It is likely that site-specific DNA binding of NKX3.1 *in vivo* is enhanced by post-translational modification and/or protein-protein interaction. For example, phosphorylation and intermolecular disulfide bond formation have been shown to increase DNA binding affinities of Nkx2.1 and Nkx2.5 (31–33). Also, protein interaction of NK-3 with the homeodomain-interacting protein kinase 2 (HIPK2) significantly increased NK-3 DNA binding affinity and transcriptional repression, independently of NK-3 phosphorylation by HIPK2 (34). In addition, Choi *et al.* have recently shown that the co-repressor protein Groucho, which by itself does not repress transcription, interacts with the homeodomain region of NK-3 to increase *in vitro* transcriptional repression (35).

There were no apparent differences in DNA binding or transcriptional repression between wild-type and R52C NKX3.1 under the described conditions. The polymorphism occurs at position 52 of NKX3.1, which is not located within the homeodomain. It has been suggested that regions outside the homeodomain may not contribute to DNA binding (25). However, it is possible that wild-type and R52C NKX3.1 differ functionally *in vivo* due to post-translational modification(s). For example, our preliminary data suggest that the polymorphism at position 52, which lies within a putative phosphorylation site at Ser48 (unpublished observation), may affect DNA binding through altered levels of phosphorylation (manuscript in preparation). Also, introduction of a cysteine at position 52 may affect redox sensitivity of DNA binding by the formation of aberrant inter- or intramolecular disulfide bonds. Further analysis may provide insights into the effects of the polymorphism on the function of NKX3.1.

## ACKNOWLEDGEMENTS

The authors would like to thank H. James Voeller for construction of the NKX3.1 expression plasmids and Cai Bowen for purification of the NKX3.1 fusion proteins. This research was supported by grants CA78327 from the NIH, DAMD17-98-1-8484 from the Department of Defense and ES-09888 from the NIEHS to E.P.G. and grant DAMD-17-99-1-9519 from the Department of Defense to D.J.S.

## REFERENCES

- Gehring, W.J., Affolter, M. and Burglin, T. (1994) *Annu. Rev. Biochem.*, **63**, 487–526.
- Kim, Y. and Nirenberg, M. (1989) *Proc. Natl Acad. Sci. USA*, **86**, 7716–7720.
- He, W.W., Sciavolino, P.J., Wing, J., Augustus, M., Hudson, P., Meissner, P.S., Curtis, R.T., Shell, B.K., Bostwick, D.G., Tindall, D.J. *et al.* (1997) *Genomics*, **43**, 69–77.
- Prescott, J.L., Blok, L. and Tindall, D.J. (1998) *Prostate*, **35**, 71–80.
- Sciavolino, P.J., Abrams, E.W., Yang, L., Austenberg, L.P., Shen, M.M. and Abate-Shen, C. (1997) *Dev. Dyn.*, **209**, 127–138.
- Bieberich, C.J., Fujita, K., He, W.W. and Jay, G. (1996) *J. Biol. Chem.*, **271**, 31779–31782.
- Kos, L., Chiang, C. and Mahon, K.A. (1998) *Mech. Dev.*, **70**, 25–34.
- Tanaka, M., Lyons, G.E. and Izumo, S. (1999) *Mech. Dev.*, **85**, 179–182.
- Bhatia-Gaur, R., Donjacour, A.A., Sciavolino, P.J., Kim, M., Desai, N., Young, P., Norton, C.R., Gridley, T., Cardiff, R.D., Cunha, G.R. *et al.* (1999) *Genes Dev.*, **13**, 966–977.
- Voeller, H.J., Augustus, M., Madlike, V., Bova, G.S., Carter, K.C. and Gelmann, E.P. (1997) *Cancer Res.*, **57**, 4455–4459.
- Lundgren, R., Kristoffersson, U., Heim, S., Mandahl, N. and Mitelman, F. (1988) *Cancer Genet. Cytogenet.*, **35**, 103–108.
- Carter, B.S., Ewing, C.M., Ward, W.S., Treiger, B.F., Aalders, T.W., Schalken, J.A., Epstein, J.I. and Isaacs, W.B. (1990) *Proc. Natl Acad. Sci. USA*, **87**, 8751–8755.
- Phillips, S.M., Morton, D.G., Lee, S.J., Wallace, D.M. and Neoptolemos, J.P. (1994) *Br. J. Urol.*, **73**, 390–395.
- Guazzi, S., Price, M., De Felice, M., Damante, G., Mattei, M.G. and Di Lauro, R. (1990) *EMBO J.*, **9**, 3631–3639.
- Chen, C.Y. and Schwartz, R.J. (1995) *J. Biol. Chem.*, **270**, 15628–15633.
- Mennerich, D., Hoffmann, S., Hadrys, T., Arnold, H.H. and Bober, E. (1999) *Biol. Chem.*, **380**, 1041–1048.
- Jorgensen, M.C., Vestergaard, P.H., Ericson, J., Madsen, O.D. and Serup, P. (1999) *FEBS Lett.*, **461**, 287–294.
- Bradford, M. (1976) *Anal. Biochem.*, **72**, 248–254.
- Amendt, B.A., Sutherland, L.B. and Russo, A.F. (1999) *J. Biol. Chem.*, **274**, 11635–11642.
- Iizumi, T., Yazaki, T., Kanoh, S., Kondo, I. and Koiso, K. (1987) *J. Urol.*, **137**, 1304–1306.
- Blackwell, T.K. and Weintraub, H. (1990) *Science*, **250**, 1104–1110.
- Tang, S. and Breitman, M.L. (1995) *Nucleic Acids Res.*, **23**, 1928–1935.
- Gong, Z. and Hew, C.L. (1994) *Biochemistry*, **33**, 15149–15158.
- Hayashi, S. and Scott, M.P. (1990) *Cell*, **63**, 883–894.
- Damante, G., Fabbro, D., Pellizzari, L., Civitareale, D., Guazzi, S., Polycarpou-Schwartz, M., Cauci, C., Quadrioglio, F., Formisano, S. and Di Lauro, R. (1994) *Nucleic Acids Res.*, **22**, 3075–3083.
- Damante, G., Pellizzari, L., Esposito, G., Fogolari, F., Viglino, P., Fabbro, D., Tell, G., Formisano, S. and Di Lauro, R. (1996) *EMBO J.*, **15**, 4992–5000.
- Blackwell, T.K., Kretzner, L., Blackwood, E.M., Eisenman, R.N. and Weintraub, H. (1990) *Science*, **250**, 1149–1151.
- Harfe, B.D. and Fire, A. (1998) *Development*, **125**, 421–429.
- Gruschus, J.M., Tsao, D.H., Wang, L.H., Nirenberg, M. and Ferretti, J.A. (1999) *J. Mol. Biol.*, **289**, 529–545.
- Civitareale, D., Lonigro, R., Sinclair, A.J. and Di Lauro, R. (1989) *EMBO J.*, **8**, 2537–2542.
- Zannini, M., Acebron, A., De Felice, M., Arnone, M.I., Martin-Perez, J., Santisteban, P. and Di Lauro, R. (1996) *J. Biol. Chem.*, **271**, 2249–2254.
- Kasahara, H. and Izumo, S. (1999) *Mol. Cell. Biol.*, **19**, 526–536.
- Arnone, M.I., Zannini, M. and Di Lauro, R. (1995) *J. Biol. Chem.*, **270**, 12048–12055.
- Kim, Y.H., Choi, C.Y., Lee, S.J., Conti, M.A. and Kim, Y. (1998) *J. Biol. Chem.*, **273**, 25875–25879.
- Choi, C.Y., Kim, Y.H., Kwon, H.J. and Kim, Y. (1999) *J. Biol. Chem.*, **274**, 33194–33197.